STAMP Security Classification Here

18 JUN 1982

EXPERT'S USE OF INFORMATION: IS IT BIASED?

RUTH H. PHELPS, PhD*
US Army Research Institute for the
Behavioral and Social Sciences
Alexandria, Virginia 22333

The quality of human information processing, judgment and decision making is rapidly becoming a critical concern for the Army due to both the development of increasingly complex weapon systems and availability of large volumes of information provided by automated sensor systems. In order to achieve operational effectiveness, the information processing capabilities of Army personnel must be comensurate with the technological sophistication of new weapon and information systems. Past research has shown, however, that even highly competent technical experts (engineers, physicians, scientists) are limited and often faulty information processors [1, 2, 3, 4, 5, 6, 7, 8]. Apparently these deficiencies are not the result of inadequate training in the specific fields of study, but are due to general limitations in human information processing skills and abilities, such as: inattention to various categories of information; inability to handle multiple sources of information, memory shortfalls; errors in drawing logical conclusions and inferences. Clearly, errors in human information processing can degrade the validity of even the expert's judgment, resulting in poor decisions that cost the Army money, soldier and materiel readiness, as well as operational effectiveness.

The purpose of this paper is to summarize two investigations of the ability of experts to logically utilize two different types of information used for Army decision making: (1) base rate (past frequency of occurrence) and (2) size of data sample (the number of cases on which the information is based). Past research has demonstrated that people in general [6] as well as some types of experts [2, 3, 5, 8] tend to ignore both base rate and sample size information when, in fact, this information is logically pertinent for making a decision or judgment. A few examples of military situations show how such biases would be costly. A strategic analyst predicts the likelihood of a Soviet invasion of Poland based only on the present configuration of

STAMP Security Classification here

PHELPS

troops and materiel, ignoring the past frequency (base rate) of this configuration when there was no invasion. A division commander, ignoring the relative unreliability of the small number of reports (size of data sample), directs an attack on Village A based on a single report indicating enemy location at Village A, while 7 out of 10 reports indicate enemy location at Village B. A systems engineer depends on the results of a single operational test, not realizing the potential unreliability of the results based on such a small sample of data.

Two specific questions about the use of base rate and sample size information by experts are addressed: (1) is the information used at all and (2) if so, how is it used (degree, consistency, pattern, etc.)? While previous research has focused on general analyses of a large number of individuals on a single problem, the methodology used in the present research emphasizes an in-depth study of the information processing for each individual on many problems. While this research used non-military problems, the results will reveal how experts, when required to solve a series of different problems (as in Army decision making), utilize base rate and sample size information. Experiment I investigated expert's use of base rate information, while Experiment II investigated their use of sample size information.[1]

EXPERIMENT I - USE OF BASE RATE INFORMATION

Method

Two groups participated in both Experiments I and II: the Expert Group was comprised of 15 PhD research psychologists having extensive training in statistics and scientific thinking employed by the US Army Research Institute and the Novice Group was comprised of 15 undergraduate students enrolled at the University of Maine. All participants were tested individually; all completed each experiment twice.

1. Details and more extensive interpretation of these two experiments, as well as other related experiments, can be found in "Use of Base Rate and Sample Size Information by Experts and Novices: Is it Biased?" by R. Phelps, J. Fetterman, C. Tolbert and S. Yachanin, obtainable from the author.

Participants were presented with the following story:

"At birth, a large percentage of babies require medical attention for any one of a variety of complications. These complications range in frequency and severity from mild and common problems to rare and serious birth defects. The detectability of these complications varies from those that are easily diagnosed, to those that are extremely difficult to diagnose. Tests can be conducted at birth to determine whether or not a baby has any of these conditions. The tests, however, vary in how well they predict the presence or absence of a given condition in any one baby.

Condition A occurs in approximately 100 of every 1000 births in the United States.

Test X has been developed to detect whether or not a child, at birth, has Condition A. The test is accurate 90% of the time, whether the baby has Condition A or not. That is, when a baby actually has Condition A the test is positive 90% of the time. When a baby does not have Condition A, the test is negative 90% of the time.

Given that the test has indicated that Condition A is present, what are the chances that a baby selected at random from all babies born in this country, in fact has Condition A?"

Two types of numerical information appear in the story: (1) the base-rate for Condition A (100/1000 in this example), and (2) the test accuracy or individuating information (90% in this example). Based on the information in the story, participants were to estimate the chances that a randomly selected baby, in fact, had Condition A.

For each problem the story remained the same but the numerical information varied. Each problem was constructed to conform to a cell in a 4 x 5 factorial design. There were four levels of base rate (.01, .10, .60, .90) and five levels of test accuracy (.01, .10, .50, .65, .90) resulting in 20 problems. There were 14 additional problems having levels different from those in the factorial; 10 of these appeared at the beginning and the remaining four were randomly mixed with the 20 experimental problems. None of these 14 problems were scored and, unknown to the participant, were included as practice and filler problems. Experimental problems appeared in the same random order for all participants. They made their probability estimates on an unmarked scale anchored at the ends by the phrases "absolutely impossible" and "absolutely certain".

PHELPS

Results

Both the relative use of the base rate and accuracy information and the information processing strategy used by each participant were assessed using three analytical measures: analysis of variance (ANOVA), W^2 (relative importance of each type of information) and an analysis of the shape of the plotted interaction of base rate and test accuracy information.

The ANOVA conducted for each participant showed that 26 of the total 30 participants used test accuracy information. However, contrary to previous findings, 26 of the 30 also used base rate information. Three single factor and 3 two factor information processing strategies were defined apriori and are summarized in Table 1. In Figure 1 example plots indicating the 3 different single factor strategies used to combine the base rate and test accuracy information are shown. In Figure 2, two strategies combining both types of information simultaneously (2 factor strategies) are shown; in the third plot (bottom panel) is shown the logically correct strategy. Beneath each plot is indicated the number of Expert and Novices using that strategy.

Four conclusions are drawn from the data contributing to Figures 1 and 2. First, most participants regardless of expertise used base rate information. Second, while all participants adopted very systematic strategies, there were at least 5 different types of strategies. Third, none adopted the logically correct strategy. Fourth, the Experts showed no differences from Novices in either the frequency in using base rate information or in their strategies.

TABLE 1

Definition of Strategies

Single Factor Strategies

1. Test Accuracy - only the test accuracy information is used, the base rate information is ignored as indicated by a main effect for accuracy but not base rate in the ANOVA. As shown in Figure 1, the plot appears as overlapping positively sloped lines.

2. Base Rate - only base rate information is used, as indicated by a significant effect for base rate but not test accuracy in the ANOVA. The plot in Figure 1 of this strategy appears as a set of flat parallel lines.

3. Interactive - only one factor is used at a time: base rates are used when test accuracy is low and accuracy is only used when base rate is low as indicated by significant main effects and interaction. The interaction plots as a converging set of lines.

Two Factor Strategies

1. Linear - both base rate and accuracy are used simultaneously but combined linearly as shown by significant main effects but no interaction The plot appears as parallel, positively sloping lines in Figure 2.

2. Multiplying - both base rate and accuracy are used simultaneously and are combined multiplicatively as shown by significant main effects and interaction. The interaction plots as a diverging fan of lines in Figure 2.

3. Logically Correct Strategy - solutions are calculated according to Bayes' theorem. The Bayesian solution plots as a barrel-shaped function as shown in the bottom panel, Figure 2.

STAMP Security Classification here

Insert last name of author(s) here PHELPS

SINGLE FACTOR STRATEGIES

ACCURACY STRATEGY

Experts: 2
Novices: 1

BASE RATE STRATEGY

Experts: 1
Novices: 2

INTERACTIVE STRATEGY

Experts: 5
Novices: 3

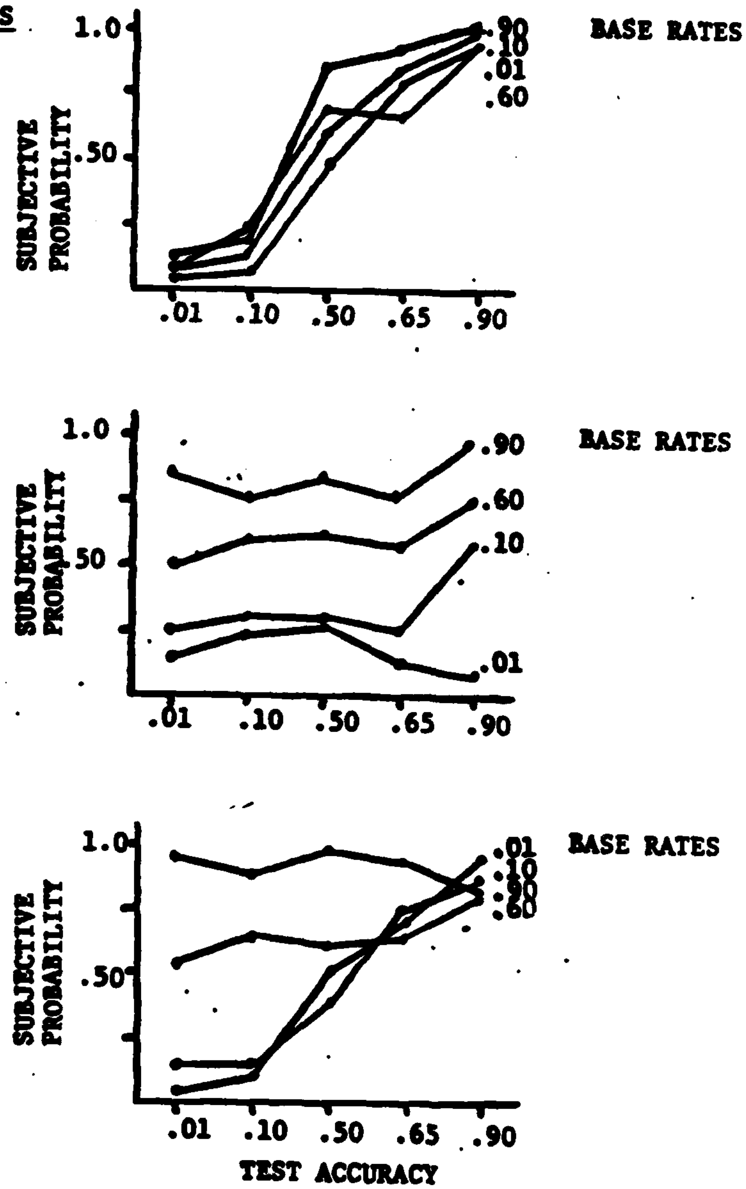

Figure 1. Plots of the Base Rate X Test Accuracy interaction for the three single factor strategies used by individual subjects. Each strategy is illustrated by data from a single subject. The number of Experts and Novices using each strategy is listed to the left of each plot.

STAMP Security Classification here

Insert last name of author(s) here PHELPS

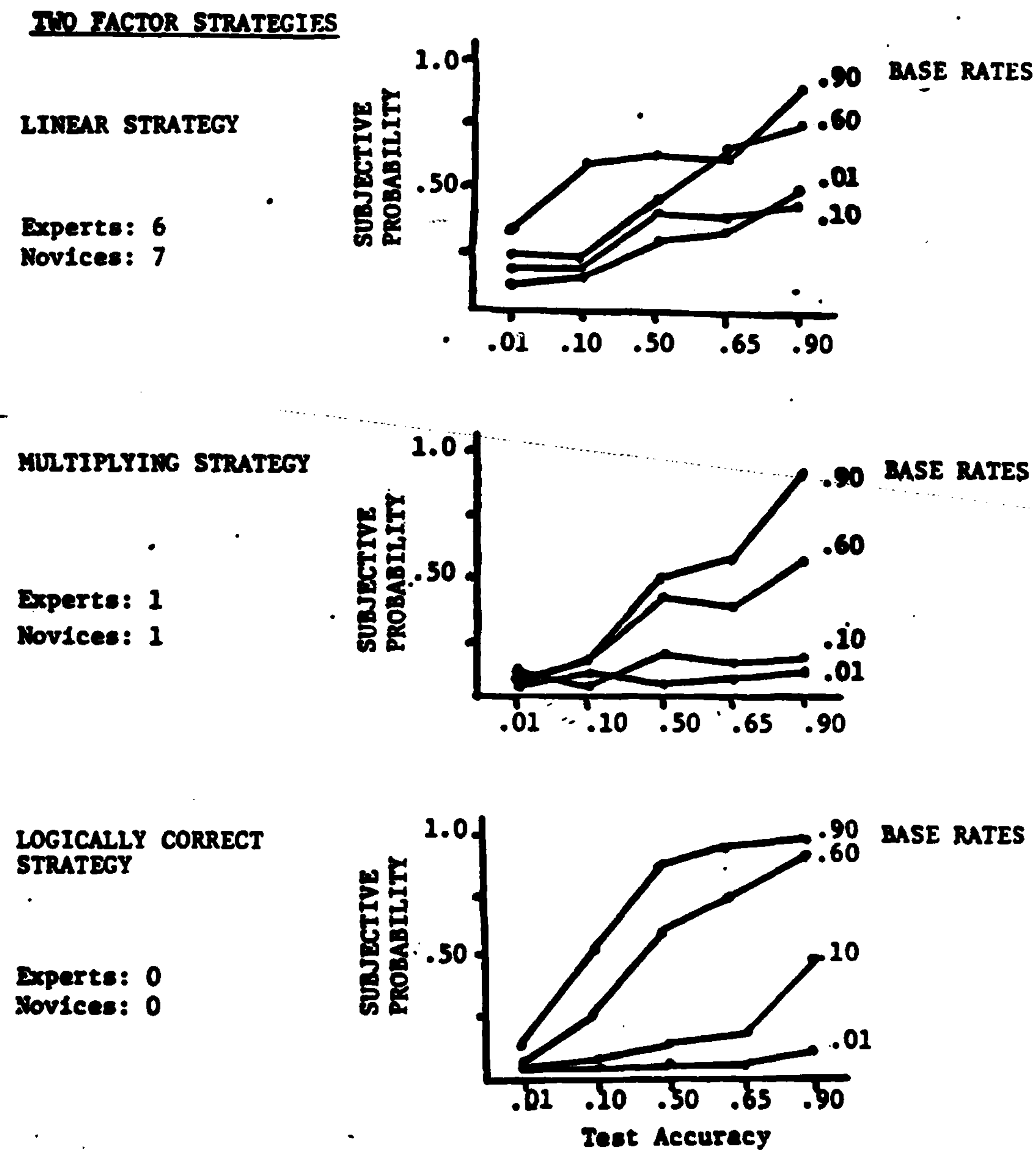

Figure 2. Plots of the Base Rate X Test Accuracy interaction for two factor strategies used by subjects (Top and Center). The bottom plot is the logically correct two factor solution derived from the application of Bayes' Theorem. The number of subjects using each strategy is shown to the left of each plot.

STAMP Security Classification here

STAMP Security Classification here

PHELPS

EXPERIMENT II-USE OF SAMPLE SIZE INFORMATION

Method

The same individuals participated in Experiment II as in Experiment I. They were presented with the following story:

> "A polling organization working for Candidate A was directed to estimate, in each of two states, the number of voters who will cast their ballots in favor of Candidate A in the fall election. In a sample of 10 voters from State X, 9 respondents said they would vote for Candidate A in the fall election. In a sample of 100 voters from State Y, 70 said they would vote for Candidate A in the upcoming election."

Two types of numerical information appear in the story: (1) the proportion of people preferring the candidate, and (2) the total number of people polled (sample size) for State X and State Y. The proportion information appeared in the story as the number preferring the candidate out of the total polled in that state (e.g., 90% in State X and 70% in State Y in the example cited above). Based on this information and the story, participants were to predict in which state the candidate had a better chance of winning and how confident they were in their prediction.

The story remained the same for all problems, while the two types of numerical information systematically varied. Each of the 81 test problems was constructed to conform to a cell in a 3x3x3x3 factorial design (4 factors, 3 levels each). The variations in factors and levels were: sample size for State X (levels: 10, 100, 500), sample size for State Y (levels: 10, 100, 500), proportion preferring the candidate in State X (levels: .01, .50, .90) and proportion preferring the candidate in State Y (levels: .01, .50, .90). There were 36 additional problems having levels different from those in the factorial; 20 of these appeared at the beginning and the remaining 16 were randomly mixed with the 81 test problems. None of these 36 problems were scored and unknown to the participants were included as practice and filler problems. Problems appeared in the same random order for all participants. They made their prediction and confidence estimates on an unmarked scale anchored at the endpoints with the phrases "Absolutely certain candidate will win in State X" and "Absolutely certain candidate will win in State Y"; the center point represented complete uncertainty.

STAMP Security Classification here

PHELPS

Results

The relative utilization of sample size and proportion preferring the candidate as well as the information processing strategy used by each individual were both assessed using three analytical measures: analysis of variance (ANOVA), W^2 (relative importance of the types of information) and an analysis of the shape of the plotted interactions of sample size and proportion information.

In contrast to the results of Experiment I, in Experiment II there were large differences between the Expert and Novice groups in their use of sample size and proportion information. While all 30 participants used proportion information, 60% (9/15) of the Experts but none of the Novices used sample size information, based on the presence of significant effects in the ANOVA and the W^2 analysis. While proportion information was uniformly more important, the sample size information nevertheless reliably accounted for an average of 20% (range: 11%-to 36%) of the systematic variance for these 9 Experts. An examination of the W^2 values for each of the 30 participants shown in Table 2 reveals the large and regular discrepancy in the use of sample size information between the Expert and Novice groups.

All participants in the Novice and 40% of the Expert group used an information processing strategy characterized by a total and exclusive reliance on the subjective difference between the proportions preferring the candidate in State X and State Y. However, 60% of the Experts used a 2-step information processing strategy: first the subjective difference between the proportions for the two states is assessed, then this difference is moderated depending on the size of the samples drawn for State X and State Y. The strategy of this moderation is when either or both of the sample sizes for State X or Y is low, winning in both states is rated equally likely; however, when both sample sizes are 100 or more, judgments are based on the difference between proportions, just as in the proportion strategy. Thus it appears that more than half of the Experts but none of the Novices appreciate that proportions based on small samples are unreliable and are therefore unwilling to assign probabilities much greater than zero.

STAMP Security Classification here

Insert last name of ...

PHELPS

Table 2

Relative Importance of Sample Size and Proportion Information (W^2 Values)

	EXPERT		NOVICE	
	Sample Size[a]	Proportion[b]	Sample Size	Proportion
01	.12	.59	0	.83
02	.08	.65	.03	.78
03	.36	.26	0	.69
04	.12	.53	0	1.0
05	.14	.24	.01	.60
06	0	.85	0	.84
07	.01	.81	0	.89
08	.18	.52	0	.83
09	.35	.27	0	.94
10	.02	.94	0	.67
11	.27	.32	.05	.61
12	.05	.31	0	.92
13	.11	.81	0	.97
14	.12	.42	0	.74
15	.04	.64	.08	.53
$\bar{X}$	.25	.53	.01	.79

a. based on the combined W^2 for sample size of State X and State Y

b. based on the W^2 for the Proportion State X and State Y interaction

STAMP Security Classification here

STAMP Security Classification here

DISCUSSION AND IMPLICATIONS

These two experiments have shown that when participants are allowed to solve several similar problems only varying in the specific numerical information: (1) both experts and novices use base rate information in contrast to previous reports; (2) there are large systematic differences in the information processing strategies; (3) none, regardless of expertise, used the logically correct information processing strategy for combining the base rate and test accuracy information; (4) only experts (60%) use sample size information, and; (5) even when sample size information is used, the strategy is extremely simplistic, not utilizing the complex implications of small unreliable samples.

Thus, these experiments show that information processing abilities of experts are better than expected for some types of problems and somewhat disappointing for others. However, the use of information is highly individualistic even within the same type of content area experts. Fortunately, the strategies used by both experts and relative novices are highly systematic and therefore amenable to scrutiny, evaluation, simulation and perhaps even modification. The problem is that complete reliance on even experts without some evaluation of the quality of the information processing could have detrimental repercussions.

On the brighter side, there are very constructive uses for the military application of these data and the methodology even though Army problems and personnel were not used in this research. At least three areas of Army concern are apparent: operational decision making, automated systems development, and training. These results as well as this methodology to study information processing could be used on those Army tasks, jobs and MOS's where information processing and integration are most critical. Then within each category of job/task, the requirements for improving the information processing abilities of personnel can be addressed through training, new systems and decision making aids.

By describing and assessing the information processing strategies and relative use of various categories of information, an evaluation of the strengths and weaknesses of Army personnel, systems, and doctrine can be made. And just as in other militarily significant skills (e.g., rifle marksmanship), the information processing skills can be targeted for additional training using the plotted strategies (as in Figures 1 and 2) as feedback to students (see [9] for a summary of the successful

use of these techniques in training experts to use relevant and not irrelevant information). In addition new automated systems may either incorporate aids to help users to be logical information processors (see [3] for example of such an aid) or even bypass the user and allow the machine to process the types of information identified as exceptionally difficult for humans [1]. Furthermore automated systems developers should further find these results to be useful in simulating the human information processing component in testing prototype systems: the natural human errors made even by experts can be incorporated into alternative system evaluations. In conclusion, the description of how both experts and non-experts process information has potentially significant payoff for Army training, systems development and operations.

SUMMARY

Two experiments were conducted to compare the use of base rate and sample size information by experts and novices in making predictions about future events. Results showed that even though the majority of experts used both types of information, they used very simplistic information processing strategies. All participants regardless of expertise used highly systematic although individualistic strategies in making their predictions. These results and methodology can be used to assess and improve Army personnel information processing through training and automated systems development.

PHELPS

REFERENCES

1. Adelman, L., Donnell, M., Phelps, R. & Patterson, J. An interactive Bayesian decision aid: Toward improving the user-aid and user-organization interfaces. Manuscript submitted for publication, 1982.

2. Elstein, A., Shulman, L. & Sprafka, S. Medical Problem Solving. Cambridge, MA: Harvard University Press, 1979.

3. Kibler, A., Watson, S., Kelly, C. & Phelps, R. A prototype decision aid for evaluating alternative courses of action for tactical engagement. (Technical Report Tr-78-A38). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences, November 1978.

4. Leaper, D., Horrocks, J., Staniland, J. & deDombal, F. Computer assisted diagnosis of abdominal pain using "estimates" provided by clinicians. British Medical Journal, 1972, 2 350-354.

5. Lusted, L. Introduction to Medical Decision Making. Springfield, IL: Thomas, 1968.

6. Nisbett, R., Borgida, E., Crandall, R. & Reed, H. Popular induction: Information is not always informative, In J.S. Caroll & J. W. Payne (Eds.) Cognition and Social Behavior. Hillsdale, NJ: Lawrence Erlbaum, Assoc., 1976.

7. Nisbett, R. & Ross, L. Human Inference: Strategies and Shortcomings of Social Judgment. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1980.

8. Tversky, A. & Kahneman, D. Judgment under uncertainty: Heuristics and biases. Science, 1974, 185, 1124-1131.

9. Shanteau, J. & Gaeth, G. Training expert decision makers to ignore irrelevant information: A comparison of lecture and interactive training procedures. (Report 81-1). Manhattan, KS: Kansas State University, 1981.